08-20-01

#6 ... 38
Prior PATENT
674538-2001
8/23

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	CHOO ET AL.
Serial No.	:	09/732,348
Filed	:	DECEMBER 7, 2000
For	:	REGULATED GENE EXPRESSION IN PLANTS

745 Fifth Avenue
New York, New York 10151

EXPRESS MAIL

Mailing Label Number: EL819166091 US

Date of Deposit: August 17, 2001

I hereby certify that this paper or fee is being deposited with the
United States Postal Service "Express Mail Post Office to
Addressee" Service under 37 CFR 1.10 on the date indicated
above and is addressed to: Assistant Commissioner for Patents,
Washington, D.C. 20231.

_____
(Typed or printed name of person mailing paper or fee)

_____
(Signature of person mailing paper or fee)

REQUEST TO ADD PRIORITY CLAIM

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

      Applicants hereby claim priority under 35 U.S.C. §119 from Great Britain

Application No. 0001580.0, filed January 24, 2000, a certified copy of which is enclosed. A new

Declaration and Power of Attorney listing this GB (or UK) patent application will be filed in due

course.

                        Respectfully submitted,

                        FROMMER LAWRENCE & HAUG LLP

By:_____
THOMAS J. KOWALSKI
Reg. No. 32,147
(212) 588-0800

00013766

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.
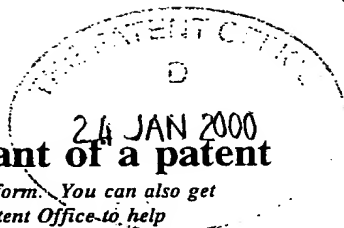
Signed

Dated   6 August 2001

Patents Form /77

nts Act 1977
(Rule 16)

The Patent Office

Request for a grant of a patent

*(See the notes on the back of this form. You can also get
an explanatory leaflet from the Patent Office to help
you fill in this form)*

24 JAN 2000

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1.	Your reference	P008355GB ATM

2.	Patent application number *(The Patent Office will fill in this part)*	**0001580.0**

TECH CENTER 1600/2900   AUG 23 2001   RECEIVED

3.	Full name, address and postcode of the or of each applicant *(underline all surnames)*	Gendaq Limited 1-3 Burtonhole Lane Mill Hill London NW7 1AD United Kingdom
	Patents ADP number *(if you know it)*	7753627001
	If the applicant is a corporate body, give the country/state of its incorporation	United Kingdom

4.	Title of the invention	Regulated Gene Expression In Plants

5.	Name of your agent *(if you have one)*	D YOUNG & CO
	"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*	21 NEW FETTER LANE LONDON EC4A 1DA
	Patents ADP number *(if you have one)*	59006

	Country	Priority application number *(if you know it)*	Date of filing *(day/month/year)*
6. If you are declaring priority from one or more earlier patent applications, give the country and date of filing of the or each of these earlier applications and (if you know it) the or each application number			

	Number of earlier application	Date of filing *(day/month/year)*
7. If this application is divided or otherwise derived from an earlier UK application, give the number and filing date of the earlier application		

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:* **YES**
   a) any applicant named in part 3 is not an inventor, or
   b) there is an inventor who is not named as an applicant, or
   c) any named applicant is a corporate body.
   See note (d))

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

   Continuation sheets of this form   0

   Description   45

   Claims *(s)*   2

   Abstract   1

   Drawing *(s)*   0

10. If you are also filing any of the following, state how many against each item.

    Priority documents

    Translations of priority documents

    Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*

    Request for preliminary examination and search *(Patents Form 9/77)*

    Request for substantive examination *(Patents Form 10/77)*

    Any other documents *(please specify)*

11.   I/We request the grant of a patent on the basis of this application.

    Signature   Date

    D YOUNG & CO
    Agents for the Applicants   24 01 00

12. Name and daytime telephone number of the person to contact in the United Kingdom   Dr A Maschio   023 8063 4816

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505

b) Write your answers in capital letters using black ink or you may type them

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

REGULATED GENE EXPRESSION IN PLANTS

Field of the Invention

5   This invention relates to the regulation of gene expression in plants using engineered zinc fingers that bind to sequences within gene regulatory sequences. Moreover, this invention also relates to transgenic plants that comprise engineered zinc fingers.

Background to the Invention

10

There has been increasing interest in the application of biotechnology to plants. For example, biotechnology has been used to improve various properties of plants such as resistance to pests. Plants also hold great promise as biological factories for a variety of chemical products including pharmaceuticals. However, genetic modifications required for

15   production of chemicals of interest are often deleterious to the plant if the corresponding gene products are continously produced. Gene switches are therefore currently of great interest to those wishing to control timing and/or dosage of gene expression in plants. Various gene switches have been developed in the prior art. In general, these prior art switches are based on naturally occurring gene transcriptional regulatory proteins.

20   However, many of these regulatory proteins have multiple gene targets since they bind sequence motifs common to the regulatory regions of a number of different genes. Furthermore, naturally occurring proteins may comprise domains that interact with endogenous molecules thus making it difficult to predict the desired outcome.

25   Summary of the Invention

The present invention seeks to overcome these difficulties by providing non-naturally occurring engineered zinc finger proteins to confer specificity on gene regulation for both transgenes and endogenous genes of interest.

30

Accordingly the present invention provides a method of regulating transcription in a plant cell from a DNA sequence comprising a target DNA operably linked to a coding sequence, which method comprises introducing an engineered zinc finger polypeptide into said plant

cell which polypeptide binds to the target DNA and modulates transcription of the coding sequence.

The term "engineered" means that the zinc finger does not occur in nature. It has therefore typically been produced by deliberate mutagenesis, for example the substitution of one or more amino acids, either as part of a random mutagenesis procedure or site-directed mutagenesis. Engineered zinc fingers for use in the invention may also have been produced de novo using rational design strategies.

The term "introduced into" means that a procedure is performed on the plant cell such that the zinc finger polypeptide is then present in the cell. Examples of suitable procedures include microinjecting presynthesised proteins or transforming/transfecting cells with a nucleic acid construct that is capable of directing expression of the zinc finger polypeptide in the cell.

In one embodiment, the target DNA is part of an endogenous genomic sequence. In another embodiment, the target DNA and coding sequence are heterologous to the cell.

The term "heterologous to the cell" means that the sequence does not naturally exist in the genome of the cell but has been introduced by whatever means, for example as part of a nucleic acid vector such as a plasmid. A heterologous sequence would preferably include a modified sequence introduced by homologous recombination such that it is present in the genome in the same position as the native allele.

In a highly preferred embodiment, the zinc finger polypeptide is fused to a biological effector domain. The term "biological effector domain" means any polypeptide that has a biological function and includes enzymes and transcriptional regulatory proteins.

Preferably the zinc finger polypeptide is fused to a transcriptional activator domain or a transcriptional repressor domain.

In a further embodiment of the method of the invention the plant cell is part of a plant and the target sequence is part of a regulatory sequence to which the nucleotide sequence of interest is operably linked.

5   The present invention further provides a plant host cell comprising a polynucleotide encoding an engineered zinc finger polypeptide and a target DNA sequence to which the zinc finger polypeptide binds.

The present invention also provides a transgenic plant comprising a polynucleotide

10   encoding an engineered zinc finger polypeptide and a target DNA sequence to which the zinc finger polypeptide binds.

Detailed Description of the Invention

15   Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art (e.g., in cell culture, molecular genetics, nucleic acid chemistry, hybridization techniques and biochemistry).

20   Standard techniques are used for molecular, genetic and biochemical methods (see generally, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. and Ausubel et al., Short Protocols in Molecular Biology (1999) 4th Ed, John Wiley & Sons, Inc. which are incorporated herein by reference), chemical methods, pharmaceutical formulations and

25   delivery and treatment of patients.

A.   Zinc fingers

A zinc finger binding motif is the α-helical structural motif found in zinc finger binding

30   proteins, well known to those skilled in the art. This is an independently folded zinc-containing mini-domain which is used in a modular repeating fashion to achieve sequence-specific recognition of DNA. The first zinc finger motif was identified in the *Xenopus*

transcription factor TFIIIA. The structure of Zf proteins has been determined by NMR studies (Lee *et al.*, 1989 Science 245, 635-637) and crystallography (Pavletich & Pabo, 1991, Science 252, 809-812).

5   Mutagenesis of Zf proteins has suggested modularity of the domains. Site directed mutagenesis has been used to change key Zf residues, identified through sequence homology alignment, and from the structural data, resulting in altered specificity of Zf domain (Nardelli *et al.*, 1992 NAR 26, 4137-4144).

10   The crystal structures of zinc finger-DNA complexes show a semiconserved pattern of interactions in which 3 amino acids from the α-helix contact 3 adjacent bases (a triplet) in DNA (Pavletich & Pabo 1991 Science 252, 809-817; Fairall *et al.*, 1993 Nature (London) 366, 483-487; and Pavletich & Pabo 1993 Science 261, 1701-1707). Thus the mode of DNA recognition is principally a one-to-one interaction between amino acids and bases. Because
15   zinc fingers function as quasi independent modules, it should be possible for fingers with different triplet specificities to be combined to give specific recognition of longer DNA sequences. Each finger is folded so that three amino acids are presented for binding to the DNA target sequence, although binding may be directly through only two of these positions. In the case of Zif268 for example, the protein is made up of three fingers which contact a 9
20   base pair contiguous sequence of target DNA. A linker sequence is found between fingers which appears to make no direct contact with the nucleic acid.

Zinc finger polypeptides according to the present invention are non-naturally occurring. That is to say, they are essentially "man-made". Typically, zinc fingers according to the invention
25   are produced by mutagenesis techniques or designed using rational design techniques. Zinc fingers may also be selected from randomised libraries using screening procedures, such as those described below.

The present invention is therefore concerned with the production of what are essentially
30   artificial or engineered DNA binding proteins. In these proteins, artificial analogues of amino acids may be used, to impart the proteins with desired properties or for other reasons. Thus, the term "amino acid", particularly in the context where "any amino acid" is referred to, means any sort of natural or artificial amino acid or amino acid analogue that

may be employed in protein construction according to methods known in the art. Moreover, any specific amino acid referred to herein may be replaced by a functional analogue thereof, particularly an artificial functional analogue. The nomenclature used herein therefore specifically comprises within its scope functional analogues or mimetics of

5   the defined amino acids.

The zinc finger polypeptide sequences to be tested and/or selected for use in the methods of the invention are typically obtained by modifying one or more amino acids residues known to be important in binding specificity. Thus, for example, zinc finger polypeptide sequences

10   may comprise a substitution at one or more of the following positions: −1, +1, +2, +3, +5 +6 and +8.

The amino acid numbering used throughout is based on the first amino acid in the α-helix of

15   the zinc finger binding motif being position +1. It will be apparent to those skilled in the art that the amino acid residue at position -1 does not, strictly speaking, form part of the α-helix of the zinc binding finger motif. Nevertheless, the residue at -1 is shown to be very important functionally and is therefore considered as part of the binding motif α-helix for the purposes of the present invention.

20

Given the lack of predictability in the outcome of rational zinc finger engineering, there is a need for a reliable method for checking the results of efforts to custom design zinc fingers with desired sequence specificity, whether such zinc fingers are obtained by design or by selection from random mutants. Not only should the target sequence be included in the test

25   assay but also related sequences because (i) selection is by affinity and not necessarily by specificity and (ii) as discussed, rational design is unreliable owing to degenerate recognition codes, incomplete code and/or unpredictable synergistic contacts.

Ideally, the assay should include all possible DNA sequences, of given length, to establish the

30   preferred specificity of the protein motif to rank other acceptable DNA sequences in terms of affinity. Therefore, wherever possible, an idea of the absolute affinity should emerge in parallel, i.e. the assay should not be simply comparative. This is possible by, for example, determining the apparent Kd of a protein for a series of related binding sites.

Zinc finger polypeptides may in one embodiment be tested individually using a plurality of DNA sequences, such as a library, as described below. For example, it may be desired to determine the preferred base recognition specificity of a zinc finger polypeptide designed
5 using rational design techniques.

In an alternative embodiment, a library of zinc finger polypeptides having different amino acids at one or more positions involved in binding specificity may be screened using an individual DNA sequence or a library of sequences and zinc finger polypeptides selected that
10 bind to a target nucleotide sequence. Such a library of sequences may conveniently be obtained by random mutagenesis at particular positions to produce a phage display library using standard techniques (see WO96/06166 for construction of a randomised Zif268 library).

15 Where a randomised zinc finger polypeptide library is used, preferably the zinc fingers are randomised at one or more of, or may have a random allocation at some or all, preferably all, of positions −1, +1, +2, +3, +5 +6, +8 and +9. More preferably, the zinc fingers are randomised at positions −1, +2, +3 and +6, and at least one of +1, +5 and +8.

20 The sequences may also be randomised at other positions (e.g. at position +9, although it is generally preferred to retain an arginine or a lysine residue at this position). Further, whilst allocation of amino acids at the designated "random" positions may be genuinely random, it is preferred to avoid a hydrophobic residue (Phe, Trp or Tyr) or a cysteine residue at such positions.

25

Preferably the zinc finger binding motif is present within the context of other amino acids (which may be present in zinc finger proteins), so as to form a zinc finger (which includes an antiparallel β-sheet). Further, the zinc finger is preferably displayed as part of a zinc finger polypeptide, which polypeptide comprises a plurality of zinc fingers joined by an intervening
30 linker peptide. Typically the library of sequences is such that the zinc finger polypeptide will comprise two or more zinc fingers of defined amino acid sequence (which may be the wild type sequence) and one zinc finger having a zinc finger binding motif randomised in the manner defined above. It is preferred that the randomised finger of the polypeptide is

positioned between the two or more fingers having defined sequence. The defined fingers will establish the "phase" of binding of the polypeptide to DNA, which helps to increase the binding specificity of the randomised finger.

5   Preferably the sequences encode the randomised binding motif of the middle finger of the Zif268 polypeptide. Conveniently, the sequences also encode those amino acids N-terminal and C-terminal of the middle finger in wild type Zif268, which encode the first and third zinc fingers respectively. In a particular embodiment, the sequence encodes the whole of the Zif268 polypeptide. Those skilled in the art will appreciate that alterations may also be made

10   to the sequence of the linker peptide and/or the β-sheet of the zinc finger polypeptide.

Typically, the randomised sequence encoding zinc finger polypeptides are such that the zinc finger binding domain can be cloned as a fusion with the minor coat protein (pIII) of bacteriophage fd. Conveniently, the encoded polypeptide includes the tripeptide sequence

15   Met-Ala-Glu as the N terminal of the zinc finger domain, which is known to allow expression and display using the bacteriophage fd system. Desirably the polypeptide library comprises 10^6 or more different sequences (ideally, as many as is practicable).

Design and testing of custom zinc fingers

20

A zinc finger binding motif is a structure well known to those in the art and defined in, for example, Miller *et al.*, (1985) EMBO J. 4:1609-1614; Berg (1988) PNAS (USA) 85:99-102; Lee *et al.*, (1989) Science 245:635-637; see WO 96/06166 and WO 96/32475, corresponding to USSN 08/422,107, incorporated herein by reference.

25

In general, a preferred zinc finger framework has the structure:

(A) $X_{0-2} \, C \, X_{1-5} \, C \, X_{9-14} \, H \, X_{3-6} \, {}^{H}/C$

30   where X is any amino acid, and the numbers in subscript indicate the possible numbers of residues represented by X.

In a preferred aspect of the present invention, zinc finger nucleic acid binding motifs may be represented as motifs having the following primary structure:

(B) X^a C X_{2-4} C X_{2-3} F X^c X X X X L X X **H** X X X^b **H** - linker

5   -1 1 2 3 4 5 6 7 8 9
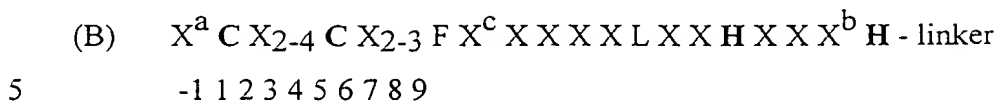
wherein X (including X^a, X^b and X^c) is any amino acid. X_{2-4} and X_{2-3} refer to the presence of 2 or 4, or 2 or 3, amino acids, respectively. The Cys and His residues, which together co-ordinate the zinc metal atom, are marked in bold text and are usually invariant,

10   as is the Leu residue at position +4 in the α-helix.

Modifications to this representation may occur or be effected without necessarily abolishing zinc finger function, by insertion, mutation or deletion of amino acids. For example it is known that the second His residue may be replaced by Cys (Krizek *et al.*,

15   (1991) J. Am. Chem. Soc. 113:4518-4523) and that Leu at +4 can in some circumstances be replaced with Arg. The Phe residue before X_c may be replaced by any aromatic other than Trp. Moreover, experiments have shown that departure from the preferred structure and residue assignments for the zinc finger are tolerated and may even prove beneficial in binding to certain nucleic acid sequences. Even taking this into account, however, the

20   general structure involving an α-helix co-ordinated by a zinc atom which contacts four Cys or His residues, does not alter. As used herein, structures (A) and (B) above are taken as an exemplary structure representing all zinc finger structures of the Cys2-His2 type.

The major binding interactions occur with amino acids -1, +3 and +6. Amino acids +4 and

25   +7 are largely invariant. The remaining amino acids may be essentially any amino acids. Preferably, position +9 is occupied by Arg or Lys. Advantageously, positions +1, +5 and +8 are not hydrophobic amino acids, that is to say are not Phe, Trp or Tyr. Preferably, position ++2 is any amino acid, and preferably serine, save where its nature is dictated by its role as a ++2 amino acid for an N-terminal zinc finger in the same nucleic acid binding

30   molecule.

In a most preferred aspect, therefore, bringing together the above, the invention allows the definition of every residue in a zinc finger DNA binding motif which will bind specifically to a given target DNA triplet.

5   The code provided by the present invention is not entirely rigid; certain choices are provided. For example, positions +1, +5 and +8 may have any amino acid allocation, whilst other positions may have certain options: for example, the present rules provide that, for binding to a central T residue, any one of Ala, Ser or Val may be used at +3. In its broadest sense, therefore, the present invention provides a very large number of proteins
10   which are capable of binding to every defined target DNA triplet.

Preferably, however, the number of possibilities may be significantly reduced. For example, the non-critical residues +1, +5 and +8 may be occupied by the residues Lys, Thr and Gln respectively as a default option. In the case of the other choices, for example, the
15   first-given option may be employed as a default. Thus, the code according to the present invention allows the design of a single, defined polypeptide (a "default" polypeptide) which will bind to its target triplet.

The present invention may be integrated with the rules set forth for zinc finger polypeptide
20   design in our copending European or PCT patent applications having publication numbers; WO 98/53057, WO 98/53060, WO 98/53058, WO 98/53059, describe improved techniques for designing zinc finger polypeptides capable of binding desired nucleic acid sequences. In combination with selection procedures, such as phage display, set forth for example in WO 96/06166, these techniques enable the production of zinc finger
25   polypeptides capable of recognising practically any desired sequence.

Verification of the results of rationally designing zinc fingers with desired specificity DNA sequences is typically carried out using a plurality of DNA sequences in addition to the sequence of interest. Libraries of sequences may conveniently be used. Typically a zinc
30   finger motif is designed as described above and then produced by recombinant or synthetic means. The zinc finger polypeptide is contacted with a DNA library and binding detected as described below. The specificity and affinity of the zinc finger for the various sequences in

the library can then be determined. If the desired binding is not seen then further modifications may be made to the zinc finger motif and the screening process repeated.

The use of automated peptide synthesisers and detection means together with computer-controlled equipment and software may allow the process to be fully automated such that when given a target sequence and rational design protocol, the process is repeated automatically until the desired result is obtained.

Screening for zinc finger polypeptides having specificity for one or more DNA sequences.

In another approach, a library of zinc finger polypeptides is contacted with a target DNA or library of DNA sequences and the zinc fingers that bind to the target sequence(s) selected. Conveniently, the zinc finger library is in the form of a library of carrier organisms that express on their surface a zinc finger polypeptide. Typical carrier organisms include phage and bacteria.

More than one round of selection may take place, for example to confirm that specificity of zinc finger polypeptides selected in any particular round. Desirably at least two, preferably three or more, rounds of screening are performed.

The library of zinc finger polypeptides need not necessarily be completely random but may be partially random, for example at certain positions only. The positions chosen and the range of different amino acids at any given position are preferably based on rational design principles.

The two methods are not mutually exclusive and may both be used as part of a design and selection strategy. For example, it may be preferred to use the screening method described above as a precursor to the rational design method described above. Thus in a preferred embodiment, that there is a two-step selection procedure: the first step comprising screening each of a plurality of zinc finger binding motifs (typically in the form of a display library), mainly or wholly on the basis of affinity for the target sequence; the second step comprising comparing binding characteristics of those motifs selected by the initial screening step, and selecting those having preferable binding characteristics for a particular DNA triplet.

The non-specific component of all protein-DNA interactions, which includes contacts to the sugar-phosphate backbone as well as ambiguous contacts to base-pairs, is a considerable driving force towards complex formation and can result in the selection of DNA-binding proteins with reasonable affinity but without specificity for a given DNA sequence. Therefore, in order to minimise these non-specific interactions when designing a polypeptide, selections should preferably be performed with low concentrations of specific binding site in a background of competitor DNA, and binding should desirably take place in solution to avoid local concentration effects and the avidity of multivalent phage for ligands immobilised on solid surfaces.

As a safeguard against spurious selections, the specificity of individual phage should be determined following the final round of selection.

B.   Target DNA

The term 'target DNA' refers to any DNA for use in the methods of the invention. This DNA may be of known sequence, or may be of unknown sequence. This DNA may be prepared artificially in a laboratory, or may be a naturally occurring DNA. This DNA may be in substantially pure form, or may be in a partially purified form, or may be part of an unpurified or heterogeneous sample. Preferably, the target DNA is a putative promoter or other transcription regulatory region such as an enhancer. More preferably, the target DNA is in substantially pure form. Even more preferably, the target DNA is of known sequence. In a most preferred embodiment, the target DNA is purified DNA of known sequence of a promoter from a plant gene of interest.

Examples of target sequences of interest include sequence motifs that are bound by transcription factors, such as zinc fingers. Particular examples include the promoters of genes involved in the biosythesis and catabolism of gibberellins (Phillips *et al.*, Plnat Physiol 108: 1049-1057 (1995), MacMillin *et al.*, Plant Physiol 113: 1369-1377 (1997), Williams *et al.*, Plant Physiol 117: 559-563 (1998); Thomas *et al.*, PNAS 96: 4698-4703 (1999)); the promoters of genes whose products are reponsible for ripening (such as polygalacturonase and ACC oxidase; the promoters of genes involved in the biosythesis of volatile ester, which are important flavour compounds in fruits and vegetables (Dudavera *et*

al., Plant Cell 8: 1137-1148 (1996); Dudavera *et al.,* Plant J. 14: 297-304 (1998); Ross *et al.,* Arch. Biochem. Biophys. 367: 9-16 (1999)); the promoters of genes involved in the biosynthesis of pharmaceutically important compounds; and the promoters of genes encoding allergens such as the peatnut allergens Arah1, Arah2 and Arah3 (Rabjohn *et al.,* J. Clin. Invest 103: 535-542).

Other plant promoters of interest are the bronze promoter (Ralston *et al.,* Genetics 119: 185-197 (1988) and Genbank Accession No. X07937.1) that directs expression of UDPglucose flavanoid glycosyl-transferase in maize, the patatin-1 gene promoter (Jefferson *et al.,* Plant Mo. Biol. 14: 995-1006 (1990)) that contains sequences capable of directing tuber-specific expression, and the phenylalanine ammonia lyase promoter (Bevan *et al.,* Embo J. 8: 1899-1906 (1989)) though to be involved in responses to mechanical wounding and normal development of the xylem and flower.

Target DNA may also be provided as a plurality of sequences, for example where one or more residues in the nucleic acid sequence are varied or random. Examples of a plurality of sequences are libraries of nucleic acid sequences comprising putative zinc finger binding sites (see below). Other sequence motifs that bind the DNA binding domain of a transcription factor may also be included in the plurality of sequences, typically varied or randomised at one or more positions. For example the chemically inducible promoter fragments described above may be randomised to produce a plurality of target DNA sequences for use in the screening methods of the present invention.

C. DNA libraries

DNA sequences for use in screening methods to select zinc fingers and corresponding DNA sequences may be provided as a library of related sequences having homology to one another (as opposed to a genomic library, for example, obtained by cutting up a large amount of essentially unrelated sequences).

A library of DNA sequences may be used in at least two different ways. Firstly, it can be used in a screen to identify zinc fingers that bind to a specific sequence. Secondly, it can be used to confirm the specificity of selected zinc fingers.

A DNA library is advantageously used to test the selectivity of a zinc finger for nucleotide sequences of length N. Consequently, since there are four different nucleotides that occur naturally in genomic DNA, the total number of sequences required to represent all possible base permutations for a sequence of length N is 4^N. N is an integer having a value of at least three. That it to say that the smallest library envisaged for testing binding to a nucleotide sequence where only one DNA triplet is varied, consists of 64 different sequences. However, N may be any integer greater than or equal to 3 such as 4, 5, 6, 7, 8 or 9. Typically, N only needs to be three times the number of zinc fingers being tested, optionally included a few additional residues outside of the binding site that may influence specificity. Thus, by way of example, to test the specificity of a protein comprising three zinc fingers, where all three fingers have been engineered, it may be desirable to use a library where N is at least 9.

Libraries of DNA sequences may be screened using a number of different methods. For example, the DNA may be immobilised to beads and incubated with zinc fingers that are labelled with an affinity ligand such as biotin or expressed on the surface of phage. Complexes between the DNA and zinc finger can be selected by washing the beads to remove unbound zinc fingers and then purifying the beads using the affinity ligand bound to the zinc fingers to remove beads that do not contain bound zinc fingers. Any remaining beads should contain DNA/zinc finger complexes. Individual beads can be selected and the identity of the DNA and zinc finger determined. Other modifications to the technique include the use of detectable labels, for example fluorescently labelling the zinc fingers and sorting beads that have zinc fingers bound to them by FACS.

In an alternative method, the DNA sequences in the library are immobilised at discrete positions on a solid substrate, such as a DNA chip, such that each different sequence is separated from other sequences on the solid substrate. Binding of zinc finger proteins is determined as described below and individual proteins isolated (which may be conveniently achieved by the use of phage display techniques). This technique may also be used as a second step after a zinc finger has been selected by, for example, the bead method described above, to characterise fully the binding specificity of a selected zinc finger

In a DNA library, it is generally not necessary or desirable for all positions to be randomised. Preferably only a subsequence of N bases of the complete DNA sequence is varied. The 4^N possible permutations of the DNA sequence of length N sequence are typically arranged in 4N sub-libraries, wherein for any one sub-library one base in the DNA sequence of length N is defined and the other N-1 bases are randomised. Thus in the case of a varied DNA triplet, there will be 12 sub-libraries.

As mentioned above, the nucleotide sequence of length N is generally part of a longer DNA molecule. However, the nucleotide sequence of length N typically occupies the same position within the longer molecule in each of the varied sequences even though the sequence of N itself may vary. The other sequences within the DNA molecule are generally the same throughout the library. Thus the library can be said to consist of a library of 4^N DNA molecules of the formula $R^1-[A/C/G/T]_4{}^N-R^2$, wherein R^1 and R^2 may be any nucleotide sequence.

Preferably, each sequence is also represented as a dilution/concentration series. Thus the immobilised DNA library may occupy $Z4^N$ discrete positions on the chip where Z is the number of different dilutions in the series and is an integer having a value of at least 2. The range of DNA concentrations for the dilution series is typically in the order of 0.01 to 100 pmol cm^{-2}, preferably from 0.05 to 5 pmol cm^{-2}. The concentrations typically vary 10-fold, i.e. a series may consist of 0.01, 0.1, 1, 10 and 100 pmol cm^{-2}, but may vary, for example, by 2- or 5-fold.

The advantage of including the DNA sequences in a dilution series is that it is then possible to estimate K_ds for protein/DNA complexes using standard techniques such as the Kaleidagraph^{TM} version 2.0 program (Abelback Software).

The DNA molecules in the library are at least partially double-stranded, in particular at least the nucleotide sequence of length N is double-stranded. Single stranded regions may be included, for example to assist in attaching the DNA library to the solid substrate.

Techniques for producing immobilised libraries of DNA molecules have been described in the art. Generally, most prior art methods described how to synthesise single-stranded

nucleic acid molecule libraries, using for example masking techniques to build up various permutations of sequences at the various discrete positions on the solid substrate. U.S. Patent No. 5,837,832, the contents of which are incorporated herein by reference, describes an improved method for producing DNA arrays immobilised to silicon substrates based on very large scale integration technology. In particular, U.S. Patent No. 5,837,832 describes a strategy called "tiling" to synthesize specific sets of probes at spatially-defined locations on a substrate which may be used to produced the immobilised DNA libraries of the present invention. U.S. Patent No. 5,837,832 also provides references for earlier techniques that may also be used.

However, an important aspect of the present invention is that it relates to DNA binding proteins, zinc fingers that bind double-stranded DNA. Thus single-stranded nucleic acid molecule libraries using the prior art techniques referred to above will then need to be converted to double-stranded DNA libraries by synthesising a complementary strand. An example of the conversion of single-stranded nucleic acid molecule libraries to double-stranded DNA libraries is given in Bulyk *et al.*, 1999, Nature Biotechnology 17, 573-577, the contents of which are incorporated herein by reference. The technique described in Bulyk *et al.*, 1999, typically requires the inclusion of a constant sequence in every member of the library (i.e. within R^1 or R^2 in the generic formula given above) to which a nucleotide primer is bound to act as a primer for second strand synthesis using a DNA polymerase and other appropriate reagents. If required, deoxynucleotide triphosphates (dNTPs) having a detectable labeled may be include to allow the efficiency of second strand synthesis to be monitored. Also the detectable label may assist in detecting binding of zinc fingers when the immobilised DNA library is in use.

Alternatively, double-stranded molecules may be synthesised off the solid substrate and each pre-formed sequence applied to a discrete position on the solid substrate. An example of such a method is to synthesis palindromic single-stranded nucleic acids – see U.S. Patent No. 5556752, the contents of which are incorporated herein by reference.

Thus DNA may typically be synthesised *in situ* on the surface of the substrate. However, DNA may also be printed directly onto the substrate using for example robotic devices equipped with either pins or pizo electric devices.

The library sequences are typically immobilised onto or in discrete regions of a solid substrate. The substrate may be porous to allow immobilisation within the substrate or substantially non-porous, in which case the library sequences are typically immobilised on the surface of the substrate. The solid substrate may be made of any material to which polypeptides can bind, either directly or indirectly. Examples of suitable solid substrates include flat glass, silicon wafers, mica, ceramics and organic polymers such as plastics, including polystyrene and polymethacrylate. It may also be possible to use semi-permeable membranes such as nitrocellulose or nylon membranes, which are widely available. The semi-permeable membranes may be mounted on a more robust solid surface such as glass. The surfaces may optionally be coated with a layer of metal, such as gold, platinum or other transition metal. A particular example of a suitable solid substrate is the commercially available BiaCore™ chip (Pharmacia Biosensors).

Preferably, the solid substrate is generally a material having a rigid or semi-rigid surface. In preferred embodiments, at least one surface of the substrate will be substantially flat, although in some embodiments it may be desirable to physically separate synthesis regions for different polymers with, for example, raised regions or etched trenches. Preferably the solid substrate is not a microtitre plate or bead. It is also preferred that the solid substrate is suitable for the high density application of DNA sequences in discrete areas of typically from 50 to 100 μm, giving a density of 10000 to 40000 cm^{-2}.

The solid substrate is conveniently divided up into sections. This may be achieved by techniques such as photoetching, or by the application of hydrophobic inks, for example teflon-based inks (Cel-line, USA).

Discrete positions, in which each different member of the library is located may have any convenient shape, e.g., circular, rectangular, elliptical, wedge-shaped, etc.

Attachment of the library sequences to the substrate may be by covalent or non-covalent means. The library sequences may be attached to the substrate via a layer of molecules to which the library sequences bind. For example, the library sequences may be labelled with biotin and the substrate coated with avidin and/or streptavidin. A convenient feature of

using biotinylated library sequences is that the efficiency of coupling to the solid substrate can be determined easily. Since the library sequences may bind only poorly to some solid substrates, it is often necessary to provide a chemical interface between the solid substrate (such as in the case of glass) and the library sequences. Examples of suitable chemical interfaces include hexaethylene glycol. Another example is the use of polylysine coated glass, the polylysine then being chemically modified using standard procedures to introduce an affinity ligand. Other methods for attaching molecules to the surfaces of solid substrate by the use of coupling agents are known in the art, see for example WO98/49557.

Binding of zinc fingers to the immobilised DNA library may be determined by a variety of means such as changes in the optical characteristics of the bound DNA (i.e. by the use of ethidium bromide) or by the use of labelled zinc finger polypeptides, such as epitope tagged zinc finger polypeptides or zinc finger polypeptides labelled with fluorophores such as green fluorescent protein. Other detection techniques that do not require the use of labels include optical techniques such as optoacoustics, reflectometry, ellipsometry and surface plasmon resonance (SPR) – see WO97/49989, incorporated herein by reference.

Binding of epitope tagged zinc finger polypeptides is typically assessed by immunological detection techniques where the primary or secondary antibody comprises a detectable label. A preferred detectable label is one that emits light, such as a fluorophore, for example phycoerythrin.

The complete DNA library is typically read at the same time by charged coupled device (CCD) camera or confocal imaging system. Alternatively, the DNA library may be placed for detection in a suitable apparatus that can move in an x-y direction, such as a plate reader. In this way, the change in characteristics for each discrete position can be measured automatically by computer controlled movement of the array to place each discrete element in turn in line with the detection means.

D.   Nucleic acid vectors encoding zinc finger proteins

Polynucleotides encoding zinc finger proteins for use in the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid

in a compatible host cell and the vector may be recovered from the host cell. Suitable host cells include bacteria such as *E. coli*, yeast and eukaryotic cell lines.

Preferably, a polynucleotide encoding a zinc finger protein according to the invention in a vector is operably linked to a control sequence that is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. The term "operably linked" means that the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under condition compatible with the control sequences.

The control sequences may be modified, for example by the addition of further transcriptional regulatory elements to make the level of transcription directed by the control sequences more responsive to transcriptional modulators.

Vectors of the invention may be transformed or transfected into a suitable host cell as described below to provide for expression of a protein of the invention. This process may comprise culturing a host cell transformed with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the protein, and optionally recovering the expressed protein.

The vectors may be for example, plasmid or virus vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a hygromycin B resistance gene for a plant vector. Vectors may be used, for example, to transfect or transform a host cell.

Control sequences operably linked to sequences encoding the protein of the invention include promoters/enhancers and other expression regulation signals such as terminators. These control sequences may be selected to be compatible with the host cell for which the expression vector is designed to be used in. The term promoter is well-known in the art and

encompasses nucleic acid regions ranging in size and complexity from minimal promoters to promoters including upstream elements and enhancers.

The promoter is typically selected from promoters which are functional in plant cells, although prokaryotic promoters and promoters functional in other eukaryotic cells may be used. The promoter is typically derived from promoter sequences of viral or plant genes. For example, it may be a promoter derived from the genome of a cell in which expression is to occur. With respect to plant promoters, they may be promoters that function in a ubiquitous manner or, alternatively, a tissue-specific manner. Tissue-specific promoters specific for different tissues of the plant are particularly preferred. Examples are provided below. Tissue-specific expression may be used to confine expression of the binding domain and/or binding partner to a cell type or tissue/organ of interest. Promoters may also be used that respond to specific stimuli, for example promoters that are responsive to plant hormones. Viral promoters may also be used, for example the CaMV 35S promoter well known in the art.

It may also be advantageous for the promoters to be inducible so that the levels of expression of the heterologous gene can be regulated during the life-time of the cell. Inducible means that the levels of expression obtained using the promoter can be regulated. Inducible expression allows the researcher to control when expression of the polypeptides takes places.

In addition, any of these promoters may be modified by the addition of further regulatory sequences, for example enhancer sequences. Chimeric promoters may also be used comprising sequence elements from two or more different promoters described above.

Many expression vectors are shuttle vectors, i.e. they are capable of replication in at least one class of organisms but can be transfected into another class of organisms for expression. For example, a vector is cloned in *E. coli* and then the same vector is transfected into yeast, mammalian or plant cells even though it is not capable of replicating independently of the host cell chromosome. DNA may also be replicated by insertion into the host genome. However, the recovery of genomic DNA encoding the zinc finger protein is more complex than that of episomally replicated vector because restriction enzyme

digestion is required to excise zinc finger protein DNA. DNA can be amplified by PCR and be directly transfected into the host cells without any replication component.

5   Advantageously, a plant expression vector encoding a zinc finger protein according to the invention may comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into host cell chromatin.

10   According to the invention, the zinc finger protein constructs of the invention are expressed in plant cells under the control of transcriptional regulatory sequences that are known to function in plants. The regulatory sequences selected will depend on the required temporal and spatial expression pattern of the zinc finger protein in the host plant. Many plant promoters have been characterized and would be suitable for use in conjunction with the invention. By way of illustration, some examples are provided below:

15

A large number of promoters are known in the art which direct expression in specific tissues and organs (*e.g.* roots, leaves, flowers) or in cell types (*e.g.* leaf epidermal cells, leaf mesophyll cells, root cortex cells). For example, the maize PEPC promoter from the phosphoenol carboxylase gene (Hudspeth & Grula Plant Mol. Bio. 12: 579-589 (1989)) is

20   green tissue-specific; the *trpA* gene promoter is pith cell-specific (WO 93/07278 to Ciba-Geigy); the TA29 promoter is pollen-specific (Mariani *et al.* Nature 347: 737-741 (1990); Mariani *et al.* Nature 357: 384-387 (1992)).

Other promoters direct transcription under conditions of presence of light or absence or

25   light or in a circadian manner. For example, the GS2 promoter described by Edwards and Coruzzi, Plant Cell 1: 241-248 (1989) is induced by light, whereas the AS1 promoter described by Tsai and Coruzzi, EMBO J 9: 323-332 (1990) is expressed only in conditions of darkness.

30   Other promoters are wound-inducible and typically direct transcription not just on wound induction, but also at the sites of pathogen infection. Examples are described by Xu *et al.* (Plant Mol. Biol. 22: 573-588 (1993)); Logemann *et al.* (Plant Cell 1: 151-158 (1989)); and Firek *et al.* (Plant Mol Biol 22: 129-142 (1993)).

A number of constitutive promoters can be used in plants. These include the Cauliflower Mosaic Virus 35S promoter (US 5,352,605 and US 5,322,938, both to Monsanto) including minimal promoters (such as the −90 CaMV 35S promoter) linked to other regulatory sequences, the rice actin promoter (McElroy *et al.* Mol. Gen. Genet. <u>231</u>: 150-160 (1991)), and the maize and sunflower ubiquitin promoters (Christensen *et al.* Plant Mol Biol. <u>12</u>: 619-632 (1989); Binet *et al.* Plant Science <u>79</u>: 87-94 (1991)).

Using promoters that direct transcription in the plant species of interest, the zinc finger protein of the invention can be expressed in the required cell or tissue types. For example, if it is the intention to utilize the zinc finger protein to regulate a gene in a specific cell or tissue type, then the appropriate promoter can be used to direct expression of the zinc finger protein construct.

An appropriate terminator of transcription is fused downstream of the selected zinc finger protein containing transgene and any of a number of available terminators can be used in conjunction with the invention. Examples of transcriptional terminator sequences that are known to function in plants include the *nopaline synthase* terminator found in the pBI vectors (Clontech catalog 1993/1994), the E9 terminator from the *rbcS* gene, and the *tml* terminator from Cauliflower Mosaic Virus.

A number of sequences found within the transcriptional unit are known to enhance gene expression and these can be used within the context of the current invention. Such sequences include intron sequences which, particularly in monocotyledonous cells, are known to enhance expression. Both intron 1 of the maize *Adh1* gene and the intron from the maize *bronze1* gene have been found to be effective in enhancing expression in maize cells (Callis *et al.* Genes Develop. <u>1</u>: 1183-1200 (1987)) and intron sequences are frequently incorporated into plant transformation vectors, typically within the non-translated leader.

A number of virus-derived non-translated leader sequences have been found to enhance expression, especially in dicotyledonous cells. Examples include the "Ω" leader sequence of Tobacco Mosaic Virus, and simlar leader sequences of Maize Chlorotic Mottle Virus

and Alfalfa Mosaic Virus (Gallie *et al.* Nucl. Acids Res. 15: 8693-8711 (1987); Shuzeski *et al.* Plant Mol Biol, 15: 65-79 (1990)).

5   The zinc finger proteins of the current invention are targeted to the cell nucleus so that they are able to interact with host cell DNA and bind to the appropriate DNA target in the nucleus and regulate transcription. To effect this, a Nuclear Localization Sequence (NLS) is incorporated in frame with the expressible zinc finger construct. The NLS can be fused either 5' or 3' to the zinc finger encoding sequence.

10   The NLS of the wild-type Simian Virus 40 Large T-Antigen (Kalderon *et al.* Cell 37: 801-813 (1984); Markland *et al.* Mol. Cell Biol. 7: 4255-4265 (1987)) is an appropriate NLS and has previously been shown to provide an effective nuclear localization mechanism in plants (van der Krol *et al.* Plant Cell 3: 667-675 (1991)). However, several alternative NLSs are known in the art and can be used instead of the SV40 NLS sequence. These

15   include the Nuclear Localization Signals of TGA-1A and TGA-1B (van der Krol *et al.*; Plant Cell 3: 667-675 (1991)).

A variety of transformation vectors are available for plant transformation and the zinc finger protein encoding genes of the invention can be used in conjunction with any such

20   vectors. The selection of vector will depend on the preferred transformation technique and the plant species which is to be transformed. For certain target species, different selectable markers may be preferred.

For *Agrobacterium*-mediated transformation, binary vectors or vectors carrying at least one

25   T-DNA border sequence are suitable. A number of vectors are available including pBIN19 (Bevan, Nucl. Acids Res. 12: 8711-8721 (1984), the pBI series of vectors, and pCIB10 and derivatives thereof (Rothstein *et al.* Gene 53: 153-161 (1987); WO 95/33818 to Ciba-Geigy).

30   Binary vector constructs prepared for *Agrobacterium* transformation are introduced into an appropriate strain of *Agrobacterium tumefaciens* (for example, LBA 4044 or GV 3101) either by triparental mating (Bevan; Nucl. Acids Res. 12: 8711-8721 (1984)) or direct transformation (Höfgen & Willmitzer, Nucl. Acids Res. 16: 9877 (1988)).

For transformation which is not *Agrobacterium*-mediated (*i.e.* direct gene transfer), any vector is suitable and linear DNA containing only the construct of interest may be preferred. Direct gene transfer can be undertaken using a single DNA species or multiple
5   DNA species (co-transformation; Schroder *et al.* Biotechnology 4: 1093-1096 (1986)).

Particularly useful for practising several embodiments of the present invention are expression vectors that provide for the transient expression of DNA encoding a zinc finger protein in plant cells. Transient expression usually involves the use of an expression vector
10   that is able to replicate efficiently in a host cell, such that the host cell accumulates many copies of the expression vector, and, in turn, synthesises high levels of zinc finger protein. For the purposes of the present invention, transient expression systems are useful e.g. for identifying zinc finger protein mutants, to identify potential phosphorylation sites, or to characterise functional domains of the protein.
15

Construction of vectors according to the invention employs conventional ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods
20   for constructing expression vectors, preparing *in vitro* transcripts, introducing DNA into host cells, and performing analyses for assessing DNA binding protein expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or
25   RNA analysis), or *in situ* hybridisation, using an appropriately labelled probe which may be based on a sequence provided herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

DNA may be stably incorporated into cells or may be transiently expressed using methods
30   known in the art. Stably transfected cells may be prepared by transfecting cells with an expression vector having a selectable marker gene, and growing the transfected cells under conditions selective for cells expressing the marker gene. To prepare transient transfectants, cells are transfected with a reporter gene to monitor transfection efficiency.

Heterologous DNA may be introduced into plant host cells by any method known in the art, such as electroporation or *Agrobacterium tumefaciens* mediated transfer. Although specific protocols may vary from species to species, transformation techniques are well

5   known in the art for most commercial plant species.

In the case of dicotyledonous species, *Agrobacterium*-mediated transformation is generally a preferred technique as it has broad application to many dicotyledons species and is generally very efficient. *Agrobacterium*-mediated transformation generally involves the

10   co-cultivation of *Agrobacterium* with explants from the plant and follows procedures and protocols that are known in the art. Transformed tissue is generally regenerated on medium carrying the appropriate selectable marker. Protocols are known in the art for many dicotyledonous crops including (for example) cotton, tomato, canola and oilseed rape, poplar, potato, sunflower, tobacco and soybean (see for example EP 0 317 511, EP 0 249

15   432, WO 87/07299, US 5,795,855).

In addition to *Agrobacterium*-mediated transformation, various other techniques can be applied to dicotyledons. These include PEG and electroporation-mediated transformation of protoplasts, and microinjection (see for example Potrykus *et al.* Mol. Gen. Genet. 199:

20   169-177 (1985); Reich *et al.* Biotechnology 4: 1001-1004 (1986); Klein *et al.* Nature 327: 70-73 (1987)). As with *Agrobacterium*-mediated transformation, transformed tissue is generally regenerated on medium carrying the appropriate selectable marker using standard techniques known in the art.

25   Although *Agrobacterium*-mediated transformation has been applied successfully to monocotyledonous species such as rice and maize and protocols for these approaches are available in the art, the most widely used transformation techniques for monocotyledons remain particle bombardment, and PEG and electroporation-mediated transformation of protoplasts.

30

In the case of maize, Gordon-Kamm *et al.* (Plant Cell 2: 603-618 (1990)), Fromm *et al.* (Biotechnology 8: 833-839 (1990) and Koziel *et al.* (Biotechnology 11: 194-200 (1993)) have published techniques for transformation using particle bombardment.

In the case of rice, protoplast-mediated transformation for both *Japonica*- and *Indica*-types has been described (Zhang *et al.* Plant Cell Rep. 7: 379-384 (1988); Shimamoto *et al.* Nature 338: 274-277; Datta *et al.* Biotechnology 8: 736-740 (1990)) and both types are also

5   routinely transformable using particle bombardment (Christou *et al.* Biotechnology 9: 957-962 (1991)).

In the case of wheat, transformation by particle bombardment has been described for both type C long-term regenerable callus (Vasil *et al.* Biotechnology 10: 667-674 (1992)) and

10   immature embryos and immature embryo-derived callus (Vasil *et al.* Biotechnology 11: 1553-1558 (1993); Weeks *et al.* Plant Physiol. 102: 1077-1084 (1993)). A further technique is described in published patent applications WO 94/13822 and WO 95/33818.

Transformation of plant cells is normally undertaken with a selectable marker which may

15   provide resistance to an antibiotic or to a herbicide. Selectable markers that are routinely used in transformation include the *nptII* gene which confers resistance to kanamycin (Messing & Vierra Gene 19: 259-268 (1982); Bevan *et al.* Nature 304: 184-187 (1983)), the *bar* gene which confers resistance to the herbicide phosphinothricin (White *et al.* Nucl. Acids Res. 18: 1062 (1990); Spencer *et al.* Theor. Appl. Genet. 79: 625-631 (1990)), the

20   *hph* gene which confers resistance to the antibiotic hygromycin (Blochlinger & Diggelmann Mol. Cell Biol. 4: 2929-2931 (1984)), and the *dhfr* gene which confers resistance to methotrexate (Bourouis *et al.* EMBO J 2: 1099-1104 (1983)). More recently, a number of selection systems have been developed which do not rely of selection for resistance to antibiotic or herbicide. These include the inducible isopentyl transferase

25   system described by Kunkel *et al.* (Nature Biotechnology 17: 916-919 (1999).

The zinc finger protein constructs of the invention are suitable for expression in a variety of different organisms. However, to enhance the efficiency of expression it may be necessary to modify the nucleotide sequence encoding the zinc finger protein to account for different

30   frequencies of codon usage in different host organisms. Hence it is preferable that the sequences to be introduced into organisms, such as plants, conform to preferred usage of codons in the host organism.

In general, high expression in plants is best achieved from codon sequences that have a GC content of at least 35% and preferably more than 45%. This is thought to be because the existence of ATTTA motifs destabilize messenger RNAs and the existence of AATAAA motifs may cause inappropriate polyadenylation, resulting in truncation of transcription. Murray *et al.* (Nucl. Acids Res. 17: 477-498 (1989)) have shown that even within plants, monocotyledonous and dicotyledonous species have differing preferences for codon usage, with monocotyledonous species generally preferring GC richer sequences. Thus, in order to achieve optimal high level expression in plants, gene sequences can be altered to accommodate such preferences in codon usage in such a manner that the codons encoded by the DNA are not changed.

Plants also have a preference for certain nucleotides adjacent to the ATG encoding the initiating methionine and for most efficient translation, these nucleotides may be modified. To facilitate translation in plant cells, it is preferable to insert, immediately upstream of the ATG representing the initiating methionine of the gene to be expressed, a "plant translational initiation context sequence". A variety of sequences can be inserted at this position. These include the sequence the sequence 5'-AAGGAGATATAACAATG-3' (Prasher *et al.* Gene 111: 229-233 (1992); Chalfie *et al.* Science 263: 802-805 (1992)), the sequence 5'-GTCGACCATG-3' (Clontech 1993/1994 catalog, page 210), and the sequence 5'-TAAACAATG-3' (Joshi *et al.* Nucl. Acids Res. 15: 6643-6653 (1987)). For any particular plant species, a survey of natural sequences available in any databank (*e.g.* GenBank) can be undertaken to determine preferred "plant translational initiation context sequences" on a species-by-species basis.

Any changes that are made to the coding sequence can be made using techniques that are well known in the art and include site directed mutagenesis, PCR, and synthetic gene construction. Such methods are described in published patent applications EP 0 385 962 (to Monsanto), EP 0 359 472 (to Lubrizol) and WO 93/07278 (to Ciba-Geigy). Well known protocols for transient expression in plants can be used to check the expression of modified genes before their transfer to plants by transformation.

E.   Regulation of gene expression *in vivo* in plants using zinc fingers

The present invention provides a method of regulating gene expression in a plant using an engineered zinc finger.

5

Thus, zinc fingers such as those designed or selected as described above are useful in switching or modulating gene expression in plants, in particular with respect to agricultural biotechnology applications as described below.

10   A fusion polypeptide comprising a zinc finger targeting domain and a DNA cleavage domain may be used to regulate gene expressing by specific cleavage of nucleic acid sequence. More usually, the zinc fingers will be fused to a transcriptional effector domain to activate or repress transcription from a gene which possesses the zinc finger binding sequence in its upstream sequences. zinc fingers capable of differentiating between U and

15   T may be used to preferentially target RNA or DNA, as required.

Thus zinc finger polypeptides according to the invention will typically require the presence of a transcriptional effector domain, such as an activation domain or a repressor domain. Examples of transcriptional activation domains include the VP16 and VP64 transactivation

20   domains of Herpes Simplex Virus. Alternative transactivation domains are various and include the maize C1 transactivation domain sequence (Sainz *et al.*, 1997, Mol. Cell. Biol. 17: 115-22) and P1 (Goff *et al.*, 1992, Genes Dev. 6: 864-75; Estruch *et al.*, 1994, Nucleic Acids Res. 22: 3983-89) and a number of other domains that have been reported from plants (see Estruch *et al.*, 1994, *ibid*).

25

Instead of incorporating a transactivator of gene expression, a repressor of gene expression can be fused to the Zinc finger protein and used to down regulate the expression of a gene contiguous or incorporating the zinc finger protein target sequence. Such repressors are known in the art and include, for example, the KRAB-A domain (Moosmann *et al.*, Biol.

30   Chem. 378: 669-677 (1997)) the *engrailed* domain (Han *et al.*, Embo J. 12: 2723-2733 (1993)) and the *snag* domain (Grimes *et al.*, Mol Cell. Biol. 16: 6263-6272 (1996)). These can be used alone or in combination to down-regulate gene expression.

Another possible application discussed above is the use of zinc fingers fused to nucleic acid cleavage moieties, such as the catalytic domain of a restriction enzyme, to produce a restriction enzyme capable of cleaving only target DNA of a specific sequence (see Kim *et al.*, (1996) Proc. Natl. Acad. Sci. USA 93:1156-1160). Using such approaches, different

5   zinc finger domains can be used to create restriction enzymes with any desired recognition nucleotide sequence. Preferably, the expression of these zinc finger-enzyme fusion proteins is inducible. It may also be possible to use enzymes other than those that cleave nucleic acids for a variety of purposes.

10   The target gene may be endogenous to the genome of the cell or may be heterologous, for example fused to a heterologous coding sequence. However, in either case it will comprise a target DNA sequence, such as a target DNA sequence described above, to which a zinc finger according to the invention binds. The zinc finger is typically expressed from a DNA construct present in the host cell comprising the target sequence. The DNA construct is

15   preferably stably integrated into the genome of the host cell, but this is not essential.

Thus a host plant cell according to the invention comprises a target DNA sequence and a construct capable of directing expression of the zinc finger molecule in the cell.

20   Suitable constructs for expressing the zinc finger molecule are known in the art and are described in section E above. The coding sequence may be expressed constitutively or be regulated. Expression may be ubiquitous or tissue-specific. Suitable regulatory sequences are known in the art and are also described in section E above. Thus the DNA construct will comprise a nucleic acid sequence encoding a zinc finger operably linked to a

25   regulatory sequence capable of directing expression of the zinc finger molecule in a host cell.

It may also be desirable to use target DNA sequences that include operably linked neighbouring sequences that bind transcriptional regulatory proteins, such as

30   transactivators. Preferably the transcriptional regulatory proteins are endogenous to the cell. If not, they will typically need to be introduced into the host cell using suitable nucleic acid constructs.

Techniques for introducing nucleic acid constructs into plant cells are known in the art and many are described both in section E and below in the section on the production of transgenic plants.

5   "Transgenic" in the present context does not encompass classical crossbreeding or *in vitro* fertilization, but rather denotes organisms in which one or more cells receive a recombinant DNA molecule. Transgenic organisms obtained by subsequent classical crossbreeding or *in vitro* fertilization of one or more transgenic organisms are included within the scope of the term "transgenic".

10

The term "germline transgenic organism" refers to a transgenic organism in which the genetic information has been taken up and incorporated into a germline cell, therefore conferring the ability to transfer the information to offspring. If such offspring, in fact, possess some or all of that information, then they, too, are transgenic multicellular

15   organisms within the scope of the present invention.

The information to be introduced into the organism is preferably foreign to the species of animal to which the recipient belongs (i.e., "heterologous"), but the information may also be foreign only to the particular individual recipient, or genetic information already

20   possessed by the recipient. In the last case, the introduced gene may be differently expressed than is the native gene.

"Operably linked" refers to polynucleotide sequences which are necessary to effect the expression of coding and non-coding sequences to which they are ligated. The nature of

25   such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence; in eukaryotes, generally, such control sequences include promoters and a transcription termination sequence. The term "control sequences" is intended to include, at a minimum, components whose presence can influence expression, and can also

30   include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

Where the nucleic acid constructs are to be integrated into the host genome, it is important to include sequences that will permit expression of polypeptides in a particular genomic context. One possible approach would to use homologous recombination to replace all or part of the endogenous gene whose expression it is desired to regulate with equivalent sequences comprising a target DNA in its regulatory sequences. This should ensure that the gene is subject to the same transcriptional regulatory mechanisms as the endogenous gene, with the exception of the target DNA sequence. Homologous recombination may also be to replace only the regulatory sequences so that the gene is subject to a different form of regulation.

In one embodiment, it is not necessary to carry out any modifications to the endogenous gene of interest since the zinc finger can be selected to bind to DNA sequences already present.

However, if the construct encoding either the zinc finger molecule or target DNA is placed randomly in the genome, it is possible that the chromatin in that region will be transcriptionally silent and in a condensed state. If this occurs, then the polypeptide will not be expressed – these are termed position-dependent effects. To overcome this problem, it may be desirable to include locus control regions (LCRs) that maintain the intervening chromatin in a transcriptionally competent open conformation. LCRs (also known as scaffold attachment regions (SARs) or matrix attachment regions (MARs)) are well known in the art – an example being the chicken lysozyme A element (Stief *et al.*, 1989, Nature 341: 343), which can be positioned around an expressible gene of interest to effect an increase in overall expression of the gene and diminish position dependent effects upon incorporation into the organism's genome (Stief *et al.*, 1989, supra). Another example is the CD2 gene LCR described by Lang *et al.*, 1991, Nucl. Acid. Res. 19: 5851-5856.

Thus, a polynucleotide construct for use in the present invention, to introduce a nucleotide sequence encoding a zinc finger molecule into the genome of a multicellular organism, typically comprises a nucleotide sequence encoding the zinc finger molecule operably linked to a regulatory sequence capable of directing expression of the coding sequence. In addition the polynucleotide construct may comprise flanking sequences homologous to the

host cell organism genome to aid in integration. An alternative approach would be to use viral vectors that are capable of integrating into the host genome, such as retroviruses.

Construction of Transgenic Plants Expressing Zinc finger Molecules

5

A transgenic plant of the invention may be produced from any plant such as the seed-bearing plants (angiosperms), and conifers. Angiosperms include dicotyledons and monocotyledons. Examples of dicotyledonous plants include tobacco, (*Nicotiana plumbaginifolia* and *Nicotiana tabacum*), arabidopsis (*Arabidopsis thaliana*), *Brassica*

10   *napus*, *Brassica nigra*, *Datura innoxia*, *Vicia narbonensis*, *Vicia faba*, pea (*Pisum sativum*), cauliflower, carnation and lentil (*Lens culinaris*). Examples of monocotyledonous plants include cereals such as wheat, barley, oats and maize.

Techniques for producing transgenic plants are well known in the art. Typically, either

15   whole plants, cells or protoplasts may be transformed with a suitable nucleic acid construct encoding a zinc finger molecule or target DNA (see above for examples of nucleic acid constructs). There are many methods for introducing transforming DNA constructs into cells, but not all are suitable for delivering DNA to plant cells. Suitable methods include *Agrobacterium* infection (see, among others, Turpen *et al.*, 1993, J. Virol. Methods, 42:

20   227-239) or direct delivery of DNA such as, for example, by PEG-mediated transformation, by electroporation or by acceleration of DNA coated particles. Acceleration methods are generally preferred and include, for example, microprojectile bombardment. A typical protocol for producing transgenic plants (in particular moncotyledons), taken from U.S. Patent No. 5, 874, 265, is described below.

25

An example of a method for delivering transforming DNA segments to plant cells is microprojectile bombardment. In this method, non-biological particles may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, gold, platinum, and the like.

30

A particular advantage of microprojectile bombardment, in addition to it being an effective means of reproducibly stably transforming both dicotyledons and monocotyledons, is that neither the isolation of protoplasts nor the susceptibility to *Agrobacterium* infection is

required. An illustrative embodiment of a method for delivering DNA into plant cells by acceleration is a Biolistics Particle Delivery System, which can be used to propel particles coated with DNA through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with plant cells cultured in suspension. The screen disperses the tungsten-

5   DNA particles so that they are not delivered to the recipient cells in large aggregates. It is believed that without a screen intervening between the projectile apparatus and the cells to be bombarded, the projectiles aggregate and may be too large for attaining a high frequency of transformation. This may be due to damage inflicted on the recipient cells by projectiles that are too large.

10

For the bombardment, cells in suspension are preferably concentrated on filters. Filters containing the cells to be bombarded are positioned at an appropriate distance below the macroprojectile stopping plate. If desired, one or more screens are also positioned between the gun and the cells to be bombarded. Through the use of techniques set forth herein one

15   may obtain up to 1000 or more clusters of cells transiently expressing a marker gene ("foci") on the bombarded filter. The number of cells in a focus which express the exogenous gene product 48 hours post-bombardment often range from 1 to 10 and average 2 to 3.

20   After effecting delivery of exogenous DNA to recipient cells by any of the methods discussed above, a preferred step is to identify the transformed cells for further culturing and plant regeneration. This step may include assaying cultures directly for a screenable trait or by exposing the bombarded cultures to a selective agent or agents.

25   An example of a screenable marker trait is the red pigment produced under the control of the R-locus in maize. This pigment may be detected by culturing cells on a solid support containing nutrient media capable of supporting growth at this stage, incubating the cells at, e.g., 18°C and greater than 180 $\mu E\ m^{-2}\ s^{-1}$, and selecting cells from colonies (visible aggregates of cells) that are pigmented. These cells may be cultured further, either in

30   suspension or on solid media.

An exemplary embodiment of methods for identifying transformed cells involves exposing the bombarded cultures to a selective agent, such as a metabolic inhibitor, an antibiotic,

herbicide or the like. Cells which have been transformed and have stably integrated a marker gene conferring resistance to the selective agent used, will grow and divide in culture. Sensitive cells will not be amenable to further culturing.

5   To use the bar-bialaphos selective system, bombarded cells on filters are resuspended in nonselective liquid medium, cultured (e.g. for one to two weeks) and transferred to filters overlaying solid medium containing from 1-3 mg/l bialaphos. While ranges of 1-3 mg/l will typically be preferred, it is proposed that ranges of 0.1-50 mg/l will find utility in the practice of the invention. The type of filter for use in bombardment is not believed to be 10   particularly crucial, and can comprise any solid, porous, inert support.

Cells that survive the exposure to the selective agent may be cultured in media that supports regeneration of plants. Tissue is maintained on a basic media with hormones for about 2-4 weeks, then transferred to media with no hormones. After 2-4 weeks, shoot 15   development will signal the time to transfer to another media.

Regeneration typically requires a progression of media whose composition has been modified to provide the appropriate nutrients and hormonal signals during sequential developmental stages from the transformed callus to the more mature plant. Developing 20   plantlets are transferred to soil, and hardened, e.g., in an environmentally controlled chamber at about 85% relative humidity, 600 ppm CO_2, and 250 μE m^{-2} s^{-1} of light. Plants are preferably matured either in a growth chamber or greenhouse. Regeneration will typically take about 3-12 weeks. During regeneration, cells are grown on solid media in tissue culture vessels. An illustrative embodiment of such a vessel is a petri dish. 25   Regenerating plants are preferably grown at about 19°C to 28°C. After the regenerating plants have reached the stage of shoot and root development, they may be transferred to a greenhouse for further growth and testing.

Genomic DNA may be isolated from callus cell lines and plants to determine the presence 30   of the exogenous gene through the use of techniques well known to those skilled in the art such as PCR and/or Southern blotting.

Several techniques exist for inserting the genetic information, the two main principles being direct introduction of the genetic information and introduction of the genetic information by use of a vector system. A review of the general techniques may be found in articles by Potrykus (Annu Rev Plant Physiol Plant Mol Biol [1991] 42:205-225) and

5   Christou (Agro-Food-Industry Hi-Tech March/April 1994 17-27).

Thus, in one aspect, the present invention relates to a vector system which carries a construct encoding a zinc finger molecule or target DNA according to the present invention and which is capable of introducing the construct into the genome of a plant.

10

The vector system may comprise one vector, but it can comprise at least two vectors. In the case of two vectors, the vector system is normally referred to as a binary vector system. Binary vector systems are described in further detail in Gynheung An et al. (1980), Binary Vectors, *Plant Molecular Biology Manual A3*, 1-19.

15

One extensively employed system for transformation of plant cells with a given promoter or nucleotide sequence or construct is based on the use of a Ti plasmid from *Agrobacterium tumefaciens* or a Ri plasmid from *Agrobacterium rhizogenes* (An et al. (1986), *Plant Physiol.* 81, 301-305 and Butcher D.N. *et al.* (1980), *Tissue Culture Methods*

20   *for Plant Pathologists,* eds.: D.S. Ingrams and J.P. Helgeson, 203-208).

Several different Ti and Ri plasmids have been constructed which are suitable for the construction of the plant or plant cell constructs described above.

25   <u>Examples of specific applications</u>

Zinc fingers according to the invention may be used to regulate the expression of a nucleotide sequence of interest in the cell of a plant. Specific applications include the following:

30

1.   Improvement of ripening characteristics in fruit. A number of genes have been identified that are involved in the ripening process (such as in ethylene biosynthesis).

Control of the ripening process via regulation of the expression of those genes will help reduce significant losses via spoilage.

2. Modification of plant growth characteristics through intervention in hormonal pathways. Many plant characteristics are controlled by hormones. Regulation of the genes involved in the production of and response to hormones will enable produce crops with altered characteristics.

3. Improvement of other characteristics by manipulation of plant gene expression. Overexpression of the Na+/H+ antiport gene has resulted in enhanced salt tolerance in Arabidopsis. Targetted zinc fingers could be used to regulate the endogenous gene.

4. Improvement of plant aroma and flavour. Pathways leading to the production of aroma and flavour compounds in vegetables and fruit are currently being elucidated allowing the enhancement of these traits using zinc finger technology.

5. Improving the pharmaceutical and nutraceutical potential of plants. Many pharmaceutically active compounds are known to exist in plants, but in many cases production is limited due to insufficient biosynthesis in plants. Zinc finger technology could be used to overcome this limitation by upregulating specific genes or biochemical pathways. Other uses include regulating the expression of genes involved in biosynthesis of commercially valuable compounds that are toxic to the development of the plant.

6. Reducing harmful plant components. Some plant components lead to adverse allergic reaction when ingested in food. Zinc finger technology could be used to overcome this problem by downregulating specific genes responsible for these reactions.

7. As well as modulating the expression of endogenous genes, heterologous genes may be introduced whose expression is regulated by zinc finger proteins. For example, a nucleotide sequence of interest may encode a gene product that is preferentially toxic to cells of the male or female organs of the plant such that the ability of the plant to reproduce can be regulated. Alternatively, or in addition, the regulatory sequences to which the nucleotide sequence is operably linked may be tissue-specific such that expression when

induced only occurs in male or female organs of the plant. Suitable sequences and/or gene products are described in WO89/10396, WO92/04454 (the TA29 promoter from tobacco) and EP-A-344,029, EP-A-412,006 and EP-A-412,911.

5   The present invention will now be described by way of the following examples, which are illustrative only and non-limiting.

EXAMPLES

10   Materials And Methods

Construction And Cloning Of Genes.
In general, procedures and materials are in accordance with guidance given in Sambrook *et al.*, Molecular Cloning. A Laboratory Manual, Cold Spring Harbor, 1989. The gene for
15   the Zif268 fingers (residues 333-420) is assembled from 8 overlapping synthetic oligonucleotides (see Choo and Klug, (1994) PNAS (USA) 91:11163-67), giving *Sfi*I and *Not*I overhangs. The genes for fingers of the phage library are synthesised from 4 oligonucleotides by directional end to end ligation using 3 short complementary linkers, and amplified by PCR from the single strand using forward and backward primers which
20   contain sites for *Not*I and *Sfi*I respectively. Backward PCR primers in addition introduce Met-Ala-Glu as the first three amino acids of the zinc finger peptides, and these are followed by the residues of the wild type or library fingers as required. Cloning overhangs are produced by digestion with *Sfi*I and *Not*I where necessary. Fragments are ligated to 1 μg similarly prepared Fd-Tet-SN vector. This is a derivative of fd-tet-DOG1
25   (Hoogenboom *et al.*, (1991) Nucleic Acids Res. 19, 4133-4137) in which a section of the pelB leader and a restriction site for the enzyme *Sfi*I (underlined) have been added by site-directed mutagenesis using the oligonucleotide:

5' CTCCTGCAGTTGGACCTGTGCCAT<u>GGCCGGCTGGGCC</u>GCATAGAATGG
30   AACAACTAAAGC 3'

which anneals in the region of the polylinker. Electrocompetent DH5α cells are transformed with recombinant vector in 200 ng aliquots, grown for 1 hour in 2xTY

medium with 1% glucose, and plated on TYE containing 15 µg/ml tetracycline and 1% glucose.

The zinc finger phage display library of the present invention contains amino acid randomisations in putative base-contacting positions from the second and third zinc fingers of the three-finger DNA binding domain of Zif268, and contains members that bind DNA of the sequence XXXXXGGCG where X is any base. Further details of the library used may be found in WO 98/53057, which is incorporated herein by reference

Example 1 - Generation of Transgenic Plants Expressing a Zinc Finger Protein Fused to a Transactivation Domain

To investigate the utility of heterologous zinc finger proteins for the regulation of plant genes, a synthetic zinc finger protein was designed and introduced into transgenic *Arabidopsis thaliana* under the control of a promoter capable of expression in a plant as described below. A second construct comprising the zinc finger protein binding sequence fused upstream of the Green Fluorescent Protein (GFP) reporter gene was also introduced into transgenic *Arabidopsis thaliana* as described in Example 2. Crossing the two transgenic lines produced progeny plants carrying both constructs in which the GFP reporter gene was expressed demonstrating transactivation of the gene by the zinc finger protein.

Using conventional cloning techniques, the following constructs were made as XbaI-BamHI fragments in the cloning vector pcDNA3.1 (Invitrogen).

pTFIIIAZifVP16

pTFIIIAZifVP16 comprises a fusion of four finger domains of the zinc finger protein TFIIIA fused to the three fingers of the zinc finger protein Zif268. The TFIIIA-derived sequence is fused in frame to the translational initiation sequence ATG. The 7 amino acid Nuclear Localization Sequence (NLS) of the wild-type Simian Virus 40 Large T-Antigen is fused to the 3' end of the Zif268 sequence, and the VP16 transactivation sequence is fused downstream of the NLS. In addition, 30 bp sequence from the *c-myc* gene is introduced

downstream of the VP16 domain as a "tag" to facilitate cellular localization studies of the trangene. While this is experimentally useful, the presence of this tag is not required for the activation (or repression) of gene expression via zinc finger proteins.

5   The sequence of pTFIIIAZifVP16 is shown in SEQ ID No. 1 as an XbaI-BamHI fragment. The translational initiating ATG is located at position 15 and is double underlined. Fingers 1 to 4 of TFIIIA extend from position 18 to position 416. Finger 4 (positions 308–416) does not bind DNA within the target sequence, but instead serves to separate the first three fingers of TFIIIA from Zif268 which is located at positions 417-689. The NLS is located

10   at positions 701-722, the VP16 transactivation domain from positions 723-956, and the c-myc tag from positions 957-986. This is followed by the translational terminator TAA.

pTFIIIAZifVP64

15   pTFIIIAZifVP64 is similar to pTFIIIAZifVP16 except that the VP64 transactivation sequence replaces the VP16 sequence of pTFIIIAZifVP16.

The sequence of pTFIIIAZifVP64 is shown in SEQ ID No. 2 as an XbaI-BamHI fragment. Locations within this sequence are as for pTFIIIAZifVP16 except that the VP64 domain is

20   located at position 723-908 and the c-myc tag from positions 909-938.

Using conventional cloning techniques, the sequence 5'-AAGGAGATATAACA-3' is introduced upstream of the translational initiating ATG of both pTFIIIAZifVP16 and pTFIIIAZifVP64. This sequence incorporates a plant translational initiation context

25   sequence to facilitate translation in plant cells (Prasher et al. Gene 111: 229-233 (1992); Chalfie et al. Science 263: 802-805 (1992)).

The final constructs are transferred to the plant binary vector pBIN121 between the Cauliflower Mosaic Virus 35S promoter and the nopaline synthase terminator sequence.

30   This transfer is effected using the XbaI site of pBIN121. The binary constructs thus derived are then introduced into Agrobacterium tumefaciens (strain LBA 4044 or GV 3101) either by triparental mating or direct transformation.

Next, *Arabidopsis thaliana* are transformed with *Agrobacterium* containing the binary vector construct using conventional transformation techniques. For example, using vacuum infiltration (*e.g.* Bechtold *et al.* CR Acad Sci Paris <u>316</u>: 1194-1199; Bent *et al.* Science <u>265</u>: 1856-1860 (1994)), transformation can be undertaken essentially as follows.

5   Seeds of *Arabidopsis* are planted on top of cheesecloth covered soil and allowed to grow at a final density of 1 per square inch under conditions of 16 hours light/8 hours dark. After 4-6 weeks, plants are ready to infiltrate. An overnight liquid culture of *Agrobacterium* carrying the appropriate construct is grown up at 28°C and used to inoculate a fresh 500ml culture. This culture is grown to an OD_{600} of at least 2.0, after which the cells are

10   harvested by centrifugation and resuspended in 1 litre of infiltration medium (1 litre prepared to contain: 2.2 g MS Salts, 1 X B5 vitamins, 50 g sucrose, 0.5 g MES pH 5.7, 0.044 μM benzylaminopurine, 200 L Silwet μL-77 (OSI Specialty)). To vacuum infiltrate, pots are inverted into the infiltration medium and placed into a vacuum oven at room temperature. Infiltration is allowed to proceed for 5 mins at 400mm Hg. After releasing

15   the vacuum, the pot is removed and layed it on its side and covered with Saran wrap. The cover is removed the next day and the plant stood upright. Seeds harvested from infiltrated plants are surface sterilized and selected on appropriate medium. Vernalizalizion is undertaken for two nights at around 4°C. Plates are then transferred to a plant growth chamber. After about 7 days, transformants are visible and are transferred to soil and

20   grown to maturity.

Transgenic plants are grown to maturity. They appear phenotypically normal and are selfed to homozygosity using standard techniques involving crossing and germination of progeny on appropriate concentration of antibiotoic.

25

Transgenic plant lines carrying the TFIIIAZifVP16 construct are designated *At*-TFIIIAZifVP16 and transgenic plant lines carrying the TFIIIAZifVP64 construct are designated *At*-TFIIIAZifVP64.

Example 2 - Generation of Transgenic Plants Carrying a Green Fluorescent Protein Reporter Gene

A reporter plasmid is constructed which incorporates the target DNA sequence of the
5 TFIIIAZifVP16 and TFIIIAZifVP64 zinc finger proteins described above upstream of the Green Fluorescent Protein (GFP) reporter gene. The target DNA sequence of TFIIIAZifVP16 and TFIIIAZifVP64 is shown in SEQ I.D. No. 3. This sequence is incorporated in single copy immediately upstream of the CaMV 35S –90 minimal promoter to which the GFP gene is fused.
10

The resultant plasmid, designated pTFIIIAZif-UAS/GFP, is transferred to the plant binary vector pBIN121 replacing the Cauliflower Mosaic Virus 35S promoter. This construct is then transferred to *Agrobacterium tumefaciens* and subsequently transferred to *Arabidopsis thaliana* as described above. Transgenic plants carrying the construct are designated *At-*
15 TFIIIAZif-UAS/GFP.

Example 3 - Use of Zinc Finger Proteins to Up-Regulate a Transgene in a Plant

To assess whether the zinc finger constructs TFIIIAZifVP16 and TFIIIAZifVP64 are able
20 to transactivate gene expression *in planta*, *Arabidopsis* lines *At*-TFIIIAZifVP16 and *At*-TFIIIAZifVP64 are crossed to *At*-TFIIIAZif-UAS/GFP. The progeny of such crosses yield plants that carry the reporter construct TFIIIAZif-UAS/GFP together with either the zinc finger protein construct TFIIIAZifVP16 or the zinc finger construct TFIIIAZifVP64.

25 Plants are screened for GFP expression using an inverted fluorescence microscope (Leitz DM-IL) fitted with a filter set (Leitz-D excitation BP 355-425, dichronic 455, emission LP 460) suitable for the main 395 nm excitation and 509 nm emission peaks of GFP.

In each case, the zinc finger construct is able to transactivate gene expression
30 demonstrating the utility of heterologous zinc finger proteins for the regulation of plant genes.

Example 4 – Generation of Transgenic Plants Expressing a Zinc Finger Fused to a Plant Transactivation domain

The constructs pTFIIIAZifVP16 and pTFIIIAZifVP64 utilize the VP16 and VP64
5 transactivation domains of Herpes Simplex Virus to activate gene expression. Alternative transactivation domains are various and include the C1 transactivation domain sequence (from maize; see Goff *et al.*; Genes Dev. 5: 298-309 (1991); Goff *et al.*; Genes Dev. 6: 864-875 (1992)), and a number of other domains that have been reported from plants (see Estruch *et al.*; Nucl. Acids Res. 22: 3983-3989 (1994)).
10

Construct pTFIIAZifC1 is made as described above for pTFIIIAZifVP16 and pTFIIIAZifVP64 except the VP16/VP64 activation domains are replaced with the C1 transactivation domain sequence

15 A transgenic *Arabidopsis* line, designated *At*-TFIIAZifC1, is produced as described above in Example 2 and crossed with *At*-TFIIIAZif-UAS/GFP. The progeny of such crosses yield plants that carry the reporter construct TFIIIAZif-UAS/GFP together with either the zinc finger protein construct TFIIIAZifC1.

20 Plants are screened for GFP expression using an inverted fluorescence microscope (Leitz DM-IL) fitted with a filter set (Leitz-D excitation BP 355-425, dichronic 455, emission LP 460) suitable for the main 395 nm excitation and 509 nm emission peaks of GFP.

Example 5 – Regulation of an endogenous plant gene – UDP glucose flavonoid
25 **glucosyl-transferase (UFGT).**

To determine whether a suitably configured zinc finger could be used to regulate gene transcription from an endogenous gene in a plant, the maize UDP glucose flavonoid glucosyl-transferase (UFGT) gene (the Bronze1 gene) was selected as the target gene.
30 UFGT is involved in anthocyanin biosynthesis. A number of wild type alleles have been identified including Bz-W22 that conditions a purple phenotypes in the maize seed and plant. The Bronze locus has been the subject of extensive genetic research because its phenotype is easy to score and its expression is tissue specific and varied (for example

aleurone, anthers, husks, cob and roots). The complete sequence of Bz-W22 including upstream regulatory sequences has been determined (Ralston *et al.*, Genetics 119: 185-197). A number of sequence motifs that bind transcriptional regulatory proteins have been identified within the Bronze promoter including sequences homologous to consensus binding sites for the myb- and myc-like proteins (Roth *et al.*, Plant Cell 3: 317-325).

Identification of a zinc finger that binds to the bronze promoter

The first step is to carry out a screen for zinc finger proteins that bind to a selected region of the Bronze promoter. A region is chosen just upstream of the AT rich block located at between -88 and -80, which has been shown to be critical for Bz1 expression (Roth *et al.*, supra).

1.   Bacterial colonies containing phage libraries that express a library of Zif268 zinc fingers randomised at one or more DNA binding residues are transferred from plates to culture medium. Bacterial cultures are grown overnight at 30°C. Culture supernatant containing phages is obtained by centrifugation.

2.   10 pmol of biotinylated target DNA, derived from the Bronze promoter, immobilised on 50 mg streptavidin beads (Dynal) is incubated with 1 ml of the bacterial culture supernatant diluted 1:1 with PBS containing 50 μM ZnCl$_2$, 4% Marvel, 2% Tween in a streptavidin coated tube for 1 hour at 20°C on a rolling platform in the presence of 4 μg poly [d(I-C)] as competitor.

3.   The tubes are washed 20 times with PBS containing 50 μM ZnCl$_2$ and 1% Tween, and 3 times with PBS containing 50 μM ZnCl$_2$ to remove non-binding phage.

4.   The remaining phage are eluted using 0.1 ml 0.1 M triethylamine and the solution is neutralised with an equal volume of 1 M Tris-Cl (pH 7.4).

5.   Logarithmic-phase *E. coli* TG1 cells are infected with eluted phage, and grown overnight, as described above, to prepare phage supernatants for subsequent rounds of selection.

6.   Single colonies of transformants obtained after four rounds of selection (steps 1 to 5) as described, are grown overnight in culture. Single-stranded DNA is prepared from phage in the culture supernatant and sequenced using the Sequenase^{TM} 2.0 kit (U.S. Biochemical Corp.). The amino acid sequences of the zinc finger clones are deduced.

Construction of a vector for expression of the zinc finger clone fused to a C1 activation domain in maize protoplasts

5   Using conventional cloning techniques and in a similar manner to Example 1, the construct pZifBz23C1 is made in cloning vector pcDNA3.1 (Invitrogen).

pZifBz23C1 comprises the three fingers of the zinc finger protein clone ZifBz23 fused in frame to the translational initiation sequence ATG. The 7 amino acid Nuclear Localization

10   Sequence (NLS) of the wild-type Simian Virus 40 Large T-Antigen is fused to the 3' end of the ZifBz23 sequence, and the C1 transactivation sequence is fused downstream of the NLS. In addition, 30 bp sequence from the *c-myc* gene is introduced downstream of the VP16 domain as a "tag" to facilitate cellular localization studies of the trangene.

15   The coding sequences of pZifBz23C1 are transferred to a plant expression vector suitable for use in maize protoplasts, the coding sequence being under the control of a constitutive CaMV 35S promoter. The resulting plasmid is termed pTMBz23. The vector also contains a hygromycin resistance gene for selection purposes.

20   A suspension culture of maize cells is prepared from calli derived from embryos obtained from inbred W22 maize stocks grown to flowering in a greenhouse and self pollinated using essentially the protocol described in EP-A-332104 (Examples 40 and 41). The suspension culture is then used to prepare protoplasts using essentially the protocol described in EP-A-332104 (Example 42).

25

Protoplasts are resuspended in 0.2 M mannitol, 0.1% w/v MES, 72 mM NaCl, 70 mM CaCl$_2$, 2.5 mM KCl, 2.5 mM glucose pH to 5.8 with KOH, at a density of about 2 x 10^6 per ml. 1 ml of the protoplast suspension is then aliquotted into plastic electroporation cuvettes and 10 µg of linearized pTMBz23 added. Electroporation is carried out s

30   described in EP-A-332104 (Example 57). Protoplasts are cultured following transformation at a density of 2 x 10^6 per ml in KM-8p medium with no solidifying agent added.

Measurements of the levels UFGT expression are made using colorimetry and/or biochemical detection methods such as Northern blots or the enzyme activity assays described by Dooner and Nelson, Proc. Natl. Acad. Sci. 74: 5623-5627 (1977). Comparison is made with mock treated protoplasts transformed with a vector only control.

Alternatively, or in addition to, analysing expression of UFGT in transformed protoplasts, intact maize plants may be recovered from transformed protoplasts and the extent of UFGT expression determined. Suitable protocols for growing up maize plants from transformed protoplasts are known in the art: Electroporated protoplasts are resuspended in Km-8p medium containing 1.2% w/v Seaplaque agarose and 1 mg/l 2,4-D. Once the gel has set, protoplasts in agarose are place in the dark at 26°C. After 14 days, clonies arise from the protoplasts. The agarose containing the colonies is transferred to the surface of a 9 cm diameter petri dish containing 30 ml of N6 medium (EP-A-332,104) containing 2,4-D solidified with 0.24% Gelrite®. 100 mg/l hygromycin B is also added to select for transformed cells. The callus is cultured further in the dark at 26°C and callus pieces subcultured every two weeks onto fresh solid medium. Pieces of callus may be analysed for the presence of the pTMBz23 construct and/or UFGT expression determined.

Corn plants are regenerated as described in Example 47 of EP-A-332,104. Plantlets appear in 4 to 8 weeks. When 2 cm tall, plantlets are transferred to ON6 medium (EP-A-332,104) in GA7 containers and roots form in 2 to 4 weeks. After transfer to peat pots plants soon become established and can then be treated as normal corn plants.

Plantlets and plants can be assayed for UFGT expression as described above.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Sequence ID 1: TFIIIA/Zif-VP16

TCTAGAGCGCCGCCATGGGAGAGAAGGCGCTGCCGGTGGTGTATAAGCGGTAC
ATCTGCTCTTTCGCCGACTGCGGCGCTGCTTATAACAAGAACTGGAAACTGCA
GGCGCATCTGTGCAAACACACAGGAGAGAAACCATTTCCATGTAAGGAAGAA
GGATGTGAGAAAGGCTTTACCTCGCTTCATCACTTAACCCGCCACTCACTCACT
CATACTGGCGAGAAAAACTTCACATGTGACTCGGATGGATGTGACTTGAGATT
TACTACAAAGGCAAACATGAAGAAGCACTTAACAGATTCCATAACATCAAGA
TCTGCGTCTATGTGTGCCATTTTGAGAACTGTGGCAAAGCATTCAAGAAACAC
AATCAATTAAAGGTTCATCAGTTCAGTCACACACAGCAGCTGCCGTATGCTTG
CCCTGTCGAGTCCTGCGATCGCCGCTTTTCTCGCTCGGATGAGCTTACCCGCCA
TATCCGCATCCACACAGGCCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTA
ACTTCAGTCGTAGTGACCACCTTACCACCCACATCCGCACCCACACAGGCGAG
AAGCCTTTTGCCTGTGACATTTGTGGGAGGAAGTTTGCCAGGAGTGATGAACG
CAAGAGGCATACCAAAATCCATTTAAGACAGAAGGACGCGGCCGCACTCGAG
CGGAATTCCGGCCCAAAAAAGAAGAGAAAGGTCGCCCCCCCGACCGATGTCA
GCCTGGGGGACGAGCTCCACTTAGACGGCGAGGACGTGGCGATGGCGCATGC
CGACGCGCTAGACGATTTCGATCTGGACATGTTGGGGGACGGGGATTCCCCGG
GGCCGGGATTTACCCCCCACGACTCCGCCCCCTACGGCGCTCTGGATACGGCC
GACTTCGAGTTTGAGCAGATGTTTACCGATGCCCTTGGAATTGACGAGTACGGT
GGGGAACAAAAACTTATTTCTGAAGAAGATCTGTAAGGATCC

Sequence ID 2: TFIIIA/Zif-VP64

TCTAGAGCGCCGCCATGGGAGAGAAGGCGCTGCCGGTGGTGTATAAGCGGTAC
ATCTGCTCTTTCGCCGACTGCGGCGCTGCTTATAACAAGAACTGGAAACTGCA
GGCGCATCTGTGCAAACACACAGGAGAGAAACCATTTCCATGTAAGGAAGAA
GGATGTGAGAAAGGCTTTACCTCGCTTCATCACTTAACCCGCCACTCACTCACT
CATACTGGCGAGAAAAACTTCACATGTGACTCGGATGGATGTGACTTGAGATT
TACTACAAAGGCAAACATGAAGAAGCACTTAACAGATTCCATAACATCAAGA
TCTGCGTCTATGTGTGCCATTTTGAGAACTGTGGCAAAGCATTCAAGAAACAC
AATCAATTAAAGGTTCATCAGTTCAGTCACACACAGCAGCTGCCGTATGCTTG ·
CCCTGTCGAGTCCTGCGATCGCCGCTTTTCTCGCTCGGATGAGCTTACCCGCCA
TATCCGCATCCACACAGGCCAGAAGCCCTTCCAGTGTCGAATCTGCATGCGTA
ACTTCAGTCGTAGTGACCACCTTACCACCCACATCCGCACCCACACAGGCGAG
AAGCCTTTTGCCTGTGACATTTGTGGGAGGAAGTTTGCCAGGAGTGATGAACG
CAAGAGGCATACCAAAATCCATTTAAGACAGAAGGACGCGGCCGCACTCGAG
CGGAATTCCGGCCCAAAAAAGAAGAGAAAGGTCGAACTTCAGCTGACTTCGG
ATGCATTAGATGACTTTGACTTAGATATGCTAGGATCTGACGCGCTAGACGATT
TCGATCTGGACATGTTGGGCAGCGATGCTCTAGACGATTTCGATTTAGATATGC
TTGGCTCGGATGCCCTGGATGACTTCGACCTCGACATGCTGTCAAGTCAGCTGA
GCCAGGAACAAAAACTTATTTCTGAAGAAGATCTGTAAGGATCC

Sequence ID 3: TFIIIA/Zif binding site

TgcgtgggcgTGTACCTggatgggagacC

CLAIMS

1. A method of regulating transcription in a plant cell from a DNA sequence comprising a target DNA operably linked to a coding sequence, which method comprises introducing an engineered zinc finger polypeptide into said plant cell which polypeptide binds to the target DNA and modulates transcription of the coding sequence.

2. A method according to claim 1 wherein the target DNA is part of an endogenous genomic sequence.

3. A method according to claim 1 wherein the target DNA and coding sequence are heterologous to the cell.

4. A method according to any one of the preceding claims wherein the zinc finger polypeptide is fused to a biological effector domain.

5. A method according to claim 4 wherein the zinc finger polypeptide is fused to a transcriptional activator domain.

6. A method according to claim 4 wherein the zinc finger polypeptide is fused to a transcriptional repressor domain.

7. A plant host cell comprising a polynucleotide encoding an engineered zinc finger polypeptide and a target DNA sequence to which the zinc finger polypeptide binds.

8. A transgenic plant comprising a polynucleotide encoding an engineered zinc finger polypeptide and a target DNA sequence to which the zinc finger polypeptide binds.

9. A method according to any one of claim 1 to 6 wherein the plant cell is part of a plant and the target sequence is part of a regulatory sequence to which the nucleotide sequence of interest is operably linked.

10.   A method according to claim 9 wherein the regulatory sequence is preferentially active in the male or female organs of the plant.

-48-

ABSTRACT

REGULATED GENE EXPRESSION IN PLANTS

A method is provided of regulating transcription in a plant cell from a DNA sequence

5   comprising a target DNA operably linked to a coding sequence, which method comprises

introducing an engineered zinc finger polypeptide in said plant cell which polypeptide

binds to the target DNA and modulates transcription of the coding sequence.